

(19)



Europäisches Patentamt  
Europäisches Patentamt  
Offices européens des brevets



(11)

**EP 0 892 056 A2**

(12)

**EUROPEAN PATENT APPLICATION**

(43) Date of publication:

20.01.1999 Bulletin 1999/03

(51) Int Cl.<sup>6</sup>: **C12N 15/31**, C07K 14/295,

C07K 16/12, A61K 39/118,

A61K 48/00, G01N 33/68,

C12Q 1/68

(21) Application number: **98305488.3**

(22) Date of filing: **10.07.1998**

(84) Designated Contracting States:

**AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU  
MC NL PT SE**

Designated Extension States:

**AL LT LV MK RO SI**

(72) Inventors:

- **Reichard, Raymond,**  
**SmithKline Beecham Pharm.**  
**King of Prussia, Pennsylvania 19406 (US)**
- **Kallender, Howard, SmithKline Beecham PLC**  
**Brentford, Middlesex TW8 9EP (GB)**

(30) Priority: **18.07.1997 US 896344**

(71) Applicants:

- **SMITHKLINE BEECHAM CORPORATION**  
**Philadelphia Pennsylvania 19103 (US)**
- **SMITHKLINE BEECHAM PLC**  
**Brentford, Middlesex TW8 9EP (GB)**

(74) Representative: **Mallalieu, Catherine Louise et al**

**D. Young & Co.,**

**21 New Fetter Lane**

**London EC4A 1DA (GB)**

(54) **GidA1 polypeptides from Chlamydia trachomatis**

(57) The invention provides gidA1 polypeptides and DNA (RNA) encoding gidA1 polypeptides and methods

for producing such polypeptides by recombinant techniques. Also provided are methods for utilizing gidA1 polypeptides to screen for antibacterial compounds.

**EP 0 892 056 A2**

## D scription

## FIELD OF THE INVENTION

5 This invention relates to newly identified polynucleotides and polypeptides, and their production and uses, as well as their variants, agonists and antagonists, and their uses. In particular, in these and in other regards, the invention relates to novel polynucleotides and polypeptides of the *gidA* family, hereinafter referred to as "gidAI".

## BACKGROUND OF THE INVENTION

10 Chlamydiaceae is a family of obligate intracellular parasites. All members share a common developmental cycle. Chlamydia infect a wide range of vertebrate host, particularly humans. Chlamydia trachomatis is one of the two recognized species of Chlamydia. Human infections caused by *Chlamydia trachomatis* are widespread. This species is one of the most common cause of sexually transmitted disease in the world. It is also one of the main causes of infertility in humans.

15 The frequency of *Chlamydia trachomatis* infections has risen dramatically in the past 20 years. This has been attributed to the emergence of multiply antibiotic resistant strains and an increasing population of people with weakened immune systems. It is no longer uncommon to isolate *Chlamydia trachomatis* strains which are resistant to some or all of the standard antibiotics. This has created a demand for both new anti-microbial agents and diagnostic tests for this organism.

20 The first described *gidA* gene was that of *E. coli* (von Meyenburg et al (1980) ICN-UCLA Symp. Mol. Cell. Biol. 19, 137-159; Swissprot accession number P17112). The closest homolog of the Chlamydia trachomatis *gidAI* polypeptide is *Pseudomonas putida* *gidA* (Ogasawara, N. & Yoshikawa, H. (1992) Mol. Microbiol. 6(5), 629-634; Swissprot accession number P25756 ). The closest homolog of the Chlamydia trachomatis *gidAI* polynucleotide is *Borrelia burgdorferi* *gidA* (Old, I.G. et al. (1992) FEMS Microbiol. Lett. 99, 245-250; Genbank accession number Z12160).

25 Clearly, there is a need for factors, such as the novel compounds of the invention, that have a present benefit of being useful to screen compounds for antibiotic activity. Such factors are also useful to determine their role in pathogenesis of infection, dysfunction and disease. There is also a need for identification and characterization of such factors and their antagonists and agonists which can play a role in preventing, ameliorating or correcting infections, dysfunctions or diseases.

30 The polypeptides of the invention have amino acid sequence homology to a known *Pseudomonas putida* *gidA* protein.

## SUMMARY OF THE INVENTION

35 It is an object of the invention to provide polypeptides that have been identified as novel *gidAI* polypeptides by homology between the amino acid sequence set out in Table 1 [SEQ ID NO: 2] and a known amino acid sequence or sequences of other proteins such as *Pseudomonas putida* *gidA* protein.

40 It is a further object of the invention to provide polynucleotides that encode *gidAI* polypeptides, particularly polynucleotides that encode the polypeptide herein designated *gidAI*.

In a particularly preferred embodiment of the invention the polynucleotide comprises a region encoding *gidAI* polypeptides comprising the sequence set out in Table 1 [SEQ ID NO: 1] which includes a full length gene, or a variant thereof.

45 In another particularly preferred embodiment of the invention there is a novel *gidAI* protein from *Chlamydia trachomatis* comprising the amino acid sequence of Table 1 [SEQ ID NO:2], or a variant thereof.

In accordance with another aspect of the invention there is provided an isolated nucleic acid molecule encoding a mature polypeptide expressible by the *Chlamydia trachomatis* D/UW-3/CX strain.

50 As a further aspect of the invention, there are provided isolated nucleic acid molecules encoding *gidAI*, particularly *Chlamydia trachomatis* *gidAI*, including mRNAs, cDNAs, genomic DNAs. Further embodiments of the invention include biologically, diagnostically, prophylactically, clinically or therapeutically useful variants thereof, and compositions comprising the same.

In accordance with another aspect of the invention, there is provided the use of a polynucleotide of the invention for therapeutic or prophylactic purposes, in particular genetic immunization. Among the particularly preferred embodiments of the invention are naturally occurring allelic variants of *gidAI* and polypeptides encoded thereby.

55 Another aspect of the invention there are provided novel polypeptides of *Chlamydia trachomatis* referred to herein as *gidAI* as well as biologically, diagnostically, prophylactically, clinically or therapeutically useful variants thereof, and compositions comprising the same.

Among the particularly preferred embodiments of the invention are variants of *gidAI* polypeptide encoded by nat-

urally occurring alleles of the gidAI gene.

In a preferred embodiment of the invention there are provided methods for producing the aforementioned gidAI polypeptides.

In accordance with yet another aspect of the invention, there are provided inhibitors to such polypeptides, useful as antibacterial agents, including, for example, antibodies.

In accordance with certain preferred embodiments of the invention, there are provided products, compositions and methods for assessing gidAI expression, treating disease, for example, classic ocular trachoma, inclusion conjunctivitis, genital trachoma, infant pneumonitis, Lymphogranuloma Venerium, incipient trachoma, keratitis, papillary hypertrophy, corneal infiltration, vulvovaginitis, ear infection, mucopurulent rhinitis, salpingitis, cervicitis, cervical follicles, prostatitis, proctitis, urethritis, lymphogranule inguinale, climatic bubo, tropical bubo, and esthiomene., assaying genetic variation, and administering a gidAI polypeptide or polynucleotide to an organism to raise an immunological response against a bacteria, especially a *Chlamydia trachomatis* bacteria.

In accordance with certain preferred embodiments of this and other aspects of the invention there are provided polynucleotides that hybridize to gidAI polynucleotide sequences, particularly under stringent conditions.

In certain preferred embodiments of the invention there are provided antibodies against gidAI polypeptides.

In other embodiments of the invention there are provided methods for identifying compounds which bind to or otherwise interact with and inhibit or activate an activity of a polypeptide or polynucleotide of the invention comprising: contacting a polypeptide or polynucleotide of the invention with a compound to be screened under conditions to permit binding to or other interaction between the compound and the polypeptide or polynucleotide to assess the binding to or other interaction with the compound, such binding or interaction being associated with a second component capable of providing a detectable signal in response to the binding or interaction of the polypeptide or polynucleotide with the compound; and determining whether the compound binds to or otherwise interacts with and activates or inhibits an activity of the polypeptide or polynucleotide by detecting the presence or absence of a signal generated from the binding or interaction of the compound with the polypeptide or polynucleotide.

In accordance with yet another aspect of the invention, there are provided gidAI agonists and antagonists, preferably bacteriostatic or bacteriocidal agonists and antagonists.

In a further aspect of the invention there are provided compositions comprising a gidAI polynucleotide or a gidAI polypeptide for administration to a cell or to a multicellular organism.

Various changes and modifications within the spirit and scope of the disclosed invention will become readily apparent to those skilled in the art from reading the following descriptions and from reading the other parts of the present disclosure.

## GLOSSARY

The following definitions are provided to facilitate understanding of certain terms used frequently herein.

"Host cell" is a cell which has been transformed or transfected, or is capable of transformation or transfection by an exogenous polynucleotide sequence.

"Identity," as known in the art, is a relationship between two or more polypeptide sequences or two or more polynucleotide sequences, as determined by comparing the sequences. In the art, "identity" also means the degree of sequence relatedness between polypeptide or polynucleotide sequences, as the case may be, as determined by the match between strings of such sequences. "Identity" and "similarity" can be readily calculated by known methods, including but not limited to those described in (*Computational Molecular Biology*, Lesk, A.M., ed., Oxford University Press, New York, 1988; *Biocomputing. Informatics and Genome Projects*, Smith, D.W., ed., Academic Press, New York, 1993; *Computer Analysis of Sequence Data*, Part I, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; *Sequence Analysis in Molecular Biology*, von Heinje, G., Academic Press, 1987; and *Sequence Analysis Primer*, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991; and Carillo, H., and Lipman, D., *SIAM J Applied Math.*, 48: 1073 (1988). Preferred methods to determine identity are designed to give the largest match between the sequences tested. Methods to determine identity and similarity are codified in publicly available computer programs. Preferred computer program methods to determine identity and similarity between two sequences include, but are not limited to, the GCG program package (Devereux, J., et al., *Nucleic Acids Research* 12(1): 387 (1984)), BLASTP, BLASTN, and FASTA (Altschul, S.F. et al., *J. Molec. Biol.* 215: 403-410 (1990). The BLAST X program is publicly available from NCB1 and other sources (*BLAST Manual*, Altschul, S., et al., NCB1 NLM NIH Bethesda, MD 20894; Altschul, S., et al., *J. Mol. Biol.* 215: 403-410 (1990). As an illustration, by a polynucleotide having a nucleotide sequence having at least, for example, 95% "identity" to a reference nucleotide sequence of SEQ ID NO: 1 it is intended that the nucleotide sequence of the polynucleotide is identical to the reference sequence except that the polynucleotide sequence may include up to five point mutations per each 100 nucleotides of the reference nucleotide sequence of SEQ ID NO: 1. In other words, to obtain a polynucleotide having a nucleotide sequence at least 95% identical to a reference nucleotide sequence, up to 5% of the nucleotides in the reference sequence may be deleted or substituted

with another nucleotide, or a number of nucleotides up to 5% of the total nucleotides in the reference sequence may be inserted into the reference sequence. These mutations of the reference sequence may occur at the 5' or 3' terminal positions of the reference nucleotide sequence or anywhere between those terminal positions, interspersed either individually among nucleotides in the reference sequence or in one or more contiguous groups within the reference sequence. Analogously, by a polypeptide having an amino acid sequence having at least, for example, 95% identity to a reference amino acid sequence of SEQ ID NO:2 is intended that the amino acid sequence of the polypeptide is identical to the reference sequence except that the polypeptide sequence may include up to five amino acid alterations per each 100 amino acids of the reference amino acid of SEQ ID NO: 2. In other words, to obtain a polypeptide having an amino acid sequence at least 95% identical to a reference amino acid sequence, up to 5% of the amino acid residues in the reference sequence may be deleted or substituted with another amino acid, or a number of amino acids up to 5% of the total amino acid residues in the reference sequence may be inserted into the reference sequence. These alterations of the reference sequence may occur at the amino or carboxy terminal positions of the reference amino acid sequence or anywhere between those terminal positions, interspersed either individually among residues in the reference sequence or in one or more contiguous groups within the reference sequence.

"Isolated" means altered "by the hand of man" from its natural state, *i.e.*, if it occurs in nature, it has been changed or removed from its original environment, or both. For example, a polynucleotide or a polypeptide naturally present in a living organism is not "isolated," but the same polynucleotide or polypeptide separated from the coexisting materials of its natural state is "isolated", as the term is employed herein.

"Polynucleotide(s)" generally refers to any polyribonucleotide or polydeoxribonucleotide, which may be unmodified RNA or DNA or modified RNA or DNA. "Polynucleotide(s)" include, without limitation, single- and double-stranded DNA, DNA that is a mixture of single- and double-stranded regions or single-, double- and triple-stranded regions, single- and double-stranded RNA, and RNA that is mixture of single- and double-stranded regions, hybrid molecules comprising DNA and RNA that may be single-stranded or, more typically, double-stranded, or triple-stranded regions, or a mixture of single- and double-stranded regions. In addition, "polynucleotide" as used herein refers to triple-stranded regions comprising RNA or DNA or both RNA and DNA. The strands in such regions may be from the same molecule or from different molecules. The regions may include all of one or more of the molecules, but more typically involve only a region of some of the molecules. One of the molecules of a triple-helical region often is an oligonucleotide. As used herein, the term "polynucleotide(s)" also includes DNAs or RNAs as described above that contain one or more modified bases. Thus, DNAs or RNAs with backbones modified for stability or for other reasons are "polynucleotide(s)" as that term is intended herein. Moreover, DNAs or RNAs comprising unusual bases, such as inosine, or modified bases, such as tritylated bases, to name just two examples, are polynucleotides as the term is used herein. It will be appreciated that a great variety of modifications have been made to DNA and RNA that serve many useful purposes known to those of skill in the art. The term "polynucleotide(s)" as it is employed herein embraces such chemically, enzymatically or metabolically modified forms of polynucleotides, as well as the chemical forms of DNA and RNA characteristic of viruses and cells, including, for example, simple and complex cells. "Polynucleotide(s)" also embraces short polynucleotides often referred to as oligonucleotide(s).

"Polypeptide(s)" refers to any peptide or protein comprising two or more amino acids joined to each other by peptide bonds or modified peptide bonds. "Polypeptide(s)" refers to both short chains, commonly referred to as peptides, oligopeptides and oligomers and to longer chains generally referred to as proteins. Polypeptides may contain amino acids other than the 20 gene encoded amino acids. "Polypeptide(s)" include those modified either by natural processes, such as processing and other post-translational modifications, but also by chemical modification techniques. Such modifications are well described in basic texts and in more detailed monographs, as well as in a voluminous research literature, and they are well known to those of skill in the art. It will be appreciated that the same type of modification may be present in the same or varying degree at several sites in a given polypeptide. Also, a given polypeptide may contain many types of modifications. Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains, and the amino or carboxyl termini. Modifications include, for example, acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphatidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cysteine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, glycosylation, lipid attachment, sulfation, gamma-carboxylation of glutamic acid residues, hydroxylation and ADP-ribosylation, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins, such as arginylation, and ubiquitination. See, for instance, *PROTEINS - STRUCTURE AND MOLECULAR PROPERTIES*, 2nd Ed., T. E. Creighton, W. H. Freeman and Company, New York (1993) and Wold, F., Posttranslational Protein Modifications: Perspectives and Prospects, pgs. 1-12 in *POSTTRANSLATIONAL COVALENT MODIFICATION OF PROTEINS*, B. C. Johnson, Ed., Academic Press, New York (1983); Seifter et al., *Meth. Enzymol.* 182:626-646 (1990) and Rattan et al., *Protein Synthesis: Posttranslational Modifications and Aging*, Ann. N.Y. Acad. Sci. 663: 48-62

(1992). Polypeptides may be branched or cyclic, with or without branching. Cyclic, branched and branched circular polypeptides may result from post-translational natural processes and may be made by entirely synthetic methods, as well.

"Variant(s)" as the term is used herein, is a polynucleotide or polypeptide that differs from a reference polynucleotide or polypeptide respectively, but retains essential properties. A typical variant of a polynucleotide differs in nucleotide sequence from another, reference polynucleotide. Changes in the nucleotide sequence of the variant may or may not alter the amino acid sequence of a polypeptide encoded by the reference polynucleotide. Nucleotide changes may result in amino acid substitutions, additions, deletions, fusions and truncations in the polypeptide encoded by the reference sequence, as discussed below. A typical variant of a polypeptide differs in amino acid sequence from another, reference polypeptide. Generally, differences are limited so that the sequences of the reference polypeptide and the variant are closely similar overall and, in many regions, identical. A variant and reference polypeptide may differ in amino acid sequence by one or more substitutions, additions, deletions in any combination. A substituted or inserted amino acid residue may or may not be one encoded by the genetic code. A variant of a polynucleotide or polypeptide may be a naturally occurring such as an allelic variant, or it may be a variant that is not known to occur naturally. Non-naturally occurring variants of polynucleotides and polypeptides may be made by mutagenesis techniques, by direct synthesis, and by other recombinant methods known to skilled artisans.

## DESCRIPTION OF THE INVENTION

The invention relates to novel *gidA1* polypeptides and polynucleotides as described in greater detail below. In particular, the invention relates to polypeptides and polynucleotides of a novel *gidA1* of *Chlamydia trachomatis*, which is related by amino acid sequence homology to *Pseudomonas putida* *gidA* polypeptide. The invention relates especially to *gidA1* having the nucleotide and amino acid sequences set out in Table 1 [SEQ ID NO: 1] and Table 1 [SEQ ID NO: 2] respectively, and to the *gidA1* nucleotide sequences of the DNA in the strain and amino acid sequences encoded thereby.

---

TABLE 1

### *gidA1* Polynucleotide and Polypeptide Sequences

(A) Sequences from *Chlamydia trachomatis* *gidA1* polynucleotide sequence [SEQ ID NO:1].

5'-agacagagctaatacacagacctttgatgagcgatttaagagaaagtgaagcattgaacaagatgctgaccaaattatgt  
ttttacttcgccgcaatattatgatcctaataaaccctggaacagcagagttgattgtggctaaaaccgccacggc  
tcattggatctgtacaattagttttgaaaaagacttcgctcgattccgaaattatgctggctgtgagttccctggata  
aacaagttattcccaaaattgacagccattgtaaacttctccatcgctggtgaaagccggcgatacactttgaga  
agaaaacctatgaggatcgggtgcttggccgttctcataggcattagctgagcttttcaaaagacttaccattattcc  
ttgcgtaaacattaagaagaaaaccatgtcatctgttaagaaaaacgaagacttaagatcgccaagcataagcgtaaa

aaaagacgtcgaagagatcggcataaaatagatagttaaattgttgggttctatgtggacgttctgttgattacga  
 tgtgatagtcattggcgaggacatgctggtgtgaagctgcgtattgtgctgctaagatgggagcgtctgtcttc  
 5 tgacctcaaatttagacactattgcaaaactcagttgtaacctgctgtaggaggtatcggtaagggaacacatcgtccga  
 gaaatcgtatgctcctgggtgggattatggctgaattacigatctatcaggaattcaattccgaatcttaatacaactaa  
 aggtcccgcagtagcgcctcgcgcicaagttgacaacagctgtatcatatgaaacgctgtttagaacagg  
 10 ttctggactgcataatcatgcaggggaacagctgaggcacttttagacaatggcgaaaaagttaggagctctacaaaa  
 gaaggttgggcctatcagggaaaaccgtgtttgtcttctgaactttatgcgaggacttattcacattggaacca  
 gaacttctctgggggacgattggcgatgcagcttcttaggtcttctgaagattaaaacgcttagggtcccattag  
 gacgtttgaaaacaggaaactcctgctcgtctgttagcctcatctattgattttctgtaatggaagagcaaccaggtgat  
 15 cataacgtttgttctgtcatcgaacgagatgtttgttctacattgccacaggtttctgtcacattactcataccac  
 tgatcaaaaaagatcttatacaaaaaatctgcacgttccgctttgtatggaggcggaatcgaagggttgggcccac  
 gatattgcccctctattgaggataaaatcgtgaaattgcagataaagaccgccaccatattttattgacctgaagga  
 20 cttaacacacaaagaagttatgtgaacggtctctctacatcgtatgccttctgatgtgcagtacgatattatcgtcgtg  
 ttctggattagaaaacgcgatcatcactcgcctgcctatgccatagaatatgattatgtcatgggaacgttattttc  
 ctctttagaatccaaactaatcgaaggactcttctgtgtgggcaaatcaatggcacaacaggatacgaagaagcggct  
 25 gcacaaggcttgattgtgagtaatgccgtaataaagtcttacgccatcctccattgttccgagccgccaagagtc  
 ttatattggggtcatgttgatgatctactaccaagtagctggacgagcctatcgcagtgtaccagtagggcagaac  
 atcgtttacttgcggcaagataacgcaggcatgagactttctactatgggcactccttaggactactgtctagcgag  
 cgctatgctatgtccaagaacaaaaagctgtatagaacaagaaaagaacgggtatctaagactttccgaaaatatgg  
 30 tgatacggtagctccattaaccaaaagtctatgtcgtccgaagttcttatcaacagcttctacagaatttctgcag  
 atgttagagatttaggtcctgtgttaggtgcttcttggagatggagatcaaatattccggctatatctccgacaacaa  
 acgttaattcgtatgtgagagatccgaaaatattcaattcctgaggatattgactaccacagcatctcagcgcttag  
 35 cttagaagctagagaaaagcttttaatttactccacgcactattgggtctgcagctagaatttctgggatttctgtgg  
 cagatattcaagtgcctatgtctcttgaagaagatgctcattaattgcgtttttgttcattgcgaagggtgcctat  
 ttttaacagctccaactcgaagaagctctctacgaacctctcacaaaatttctgtctggttaatacacatcttccgg  
 aagctgtgtattgggtatttccgcaagcctgaacgagatctcacgtagaacattaaaagaagacggaatccctatt  
 40 atccgtcgtatagcggcgaggaggtgttttttggatgcagatagctgtatggtctcctggattataaattcccctac  
 tccatctccatcttcaaaagatctcctgcaatggactcaagatattatgctcctatcttccctacaggatttaaaatta  
 cagaaaatgattacactttttggataaaaaaattggcggaatgcgcagtacatccaaaatatcgtgggttcacac  
 45 acaacatttcttgaatatgaatcctaaaaagctgtcgtcctcctactcctgaaattcaaccttctatcgaca  
 aaatcgttctcatgatgaatttctaaccacaatctatgaactttttgattcaagagaagatttctatccagttaaagc  
 aatctgcagcaagtaagatgggtgtgggaacaaggttctat-3'

(B) gidA1 polypeptide sequence deduced from the polynucleotide sequence in this table  
 [SEQ ID NO:2].

NH<sub>2</sub>-

MWTFPVDYDVIVIGAGHAGCEAAYCAAKMGASVLLLTSLNLDTIAKLSCNPAVGGIGKGHIVREIDALGGI  
5 MAEITDLSGI  
QFRILNQTKGPVAVRAPRAQVDKQLYHIHMKRLLEQVPGLHIMQGTAEALLDNGEKVLGVSTKEGWAYLGK  
TVVLSSGTFM  
RGLIHIGTQNFSGGRLGDAASLGLSEDLKRLGFPLGRLKTGTPARLLASSIDFSVMEEQPGDHNVCVHR  
10 NEMFVPTLPQ  
VSCHITHTTDQTKDLITKNLHRSALYGGRIEYGVPYCPYCSIEDKIVKFADKDRHHIFIEPEGLNTQEVYV  
NGLSTSMFPD  
VQYDIIRSVSGLENAIITRPAYAIEYDYVHGNVIFPSLESKLIEGLFLCGQINGTTGYEEAAQGLIAGV  
15 NAVNKVLRHP  
PFVPSRQESYIGVMLDDLTTQVLDEPYRMFTSRAEHRLLLRQDNAGMRLSHYGHSLGLLSSERYAMFQEQ  
KACIEQEKER  
LSKTRFRKYGDTVPLTKVLCRPEVSYQQLLTEFPADVRDLGPVVGASLEMEIKYSGYISRQOTLIRSMER  
20 SENISIPEDI  
DYHSISALSLEAREKLSKFTPTIGSAARISGISVADIQVLMVSLKKDAH-COOH

(C) Polynucleotide sequence embodiments [SEQ ID NO:1].

X-(R<sub>1</sub>)<sub>n</sub>-

agacagagctaatacacagacctttgatgagcgatttaagagaaaagtggaagcattgaacaagatgctgac  
caaattatgt  
30 ttttacttcgccggaatattatgatcctaataataaacctggaacagcagagttgattgtggctaaaaa  
ccgccacggc  
tccattggatctgtacaattagtttttgaaaaagacttcgctcgattccgaaattatgctggctgtgagt  
tccctggata  
35 aacaagtttattcccaaaaattgacagccattgtaaaacttcttcccatcgctggtgaaagccggcgata  
cactttgaga  
agaaaacctatgaggatcgggtgcttgcccggttctcataggcattagctgagtttttcaaagacttac  
caattattcc  
40 ttgcgtaaaccattaagaagaaaaccatgtcatctgttaagaaaaaacgaagacttaagatcgccaagca  
taagcgtaaa  
aaaagacgtcgaagagatcggcataaaaatagatagttaaattgtttgggttctatgtggacgtttcctg  
ttgattacga  
45 tgtgatagtcattggcgcaggacatgctggttgtgaagctgcgtatttgtgctgctaagatgggagcgtct  
gtcttgcttc  
tgacctcaaatttagacactattgcaaaactcagttgtaatcctgctgtaggaggtatcggttaaggga  
catcgtccga  
50 gaaatcgatgctctcggtgggattatggctgaaattactgatctatcaggaattcaattccgaatcttaa  
atcaaactaa

55

agggtcccgagtacgcgctcctcgcgctcaagttgacaaacagctgtatcatattcatatgaaacgcttg  
ttagaacagg  
5 ttcctggactgcatatcatgcagggaaacagctgaggcacttttagacaatggcgaaaaagtttaggagt  
ctctacaaaa  
gaagggtgggcctatctagggaaaaccggtgttttgtcttctggaacttttatgcgaggacttattcaca  
ttggaacca  
10 gaacttctctgggggacgattggggcgatgcagcttctttaggtcttctgaagatttaaacgcttaggg  
ttcccaattag  
gacgtttgaaaacaggaactcctgctcgtctgttagcctcatctattgatttttctgtaatggaagagca  
accaggtgat  
15 cataacgtttgttctggtcatcgaaacgagatgtttgttctacattgccacagggttcttgtcacatta  
ctcataccac  
tgatcaaacaaaagatcttataaccaaaaatctgcatcggtccgcttgtatggagggcgaaatcgaaggg  
gttgggccac  
20 gatattgcccctctattgaggataaaaatcgtgaaatttgcagataaagaccgccaccatatttttattga  
gcctgaagga  
cttaacacacaagaagtttatgtgaacggtctctctacatcgatgcctttcgatgtgcagtacgatatta  
ttcgctcggt  
25 ttctggattagaaaacgcgatcatcactcgccctgcctatgccatagaatatgattatgttcatgggaac  
gttatttttc  
cttctttagaatccaaactaatcgaaggactcttcttgtgtgggcaaataatggcacaacaggatacga  
agaagcggct  
30 gcacaaggcttgattgctggagttaatgccgtaaataaagtcttacgccatcctccatttgttccgagcc  
gccaagagtc  
ttatatgtggggtcatgttggtgatctcactaccaagtactggacgagccttatcgcatgttcaccagt  
agggcagaac  
35 atcgtttattactgcggaagataacgcaggcatgagactttctcactatgggcactccttaggactact  
gtctagcgag  
cgctatgctatgttccaagaacaaaaagcttgtatagaacaagaaaaagaacggttatctaagactttcc  
gaaaatatgg  
40 tgatacggtagtcccattaaccaaagttctatgtcgtcccgaagtttcttatcaacagcttcttacagaa  
tttctctgcag  
atgttagagatttaggtcctgttgttaggtgcttcttggagatggagatcaaataattccggctatatctc  
ccgacaacaa  
45 acgttaattcgtatgtaggatccgaaaatatctcaattcctgaggatattgactaccacagcatct  
cagcgcttag  
cttagaagctagagaaaagctttctaaatttactccacgcactattgggtctgcagctagaatttctggg  
atttctgtgg  
50 cagatattcaagtgtcatggtctctttgaagaaagatgctcattaattgcgtttttgttcattgcgaag  
ggctgcctat



ttttaaacagctccaactcgaagaagctcttctacgaacctcttcacaaaatttctgtctggttaaataca  
 catcttccgg  
 5 aagctgtggtattgggtatttcccgaagcctgaacgagatcttcacgtagaacattttaaagaagacgg  
 aatccctatt  
 atccgtcgctatagcggcggagggactgttttttggatgcagatagcttgatggtctcctggattataa  
 attcccctac  
 10 tccatctccatcttcaaaagatctcctgcaatggactcaagatatattatgctcctatcttccctacagga  
 tttaaaatta  
 cagaaaatgattacacttttttggataaaaaaattggcggaatgcgcagtagatccaaaaatatcgatg  
 ggttcatcac  
 15 acaacatttctttggaatatgaatcctaaaaagcttgctcgctaccttctactcctgaaattcaacctt  
 cttatcgaca  
 aaatcgttctcatgatgaatttctaaccacaatctatgaactttttgattcaagagaagatttcttatcc  
 cagttaaagc  
 20 aatctgcagcaagtaagatggtgtggaacaaggttctat - (R<sub>2</sub>)<sub>n</sub>-Y

(D) Polypeptide sequence embodiments [SEQ ID NO:2].

25 X - (R<sub>1</sub>)<sub>n</sub>-  
 MWTFPVDYDVIVIGAGHAGCEAAYCAAKMGASVLLLTSLNLDTIAKLSCNPAVGGIGKGHIVREIDALGGI  
 MAEITDLSGI  
 QFRILNQTKGPAVRAPRAQVDKQLYHIHMKRLLEQVPLHIMQGTAEALLDNGEKVLGVSTKEGWAYLGK  
 30 TVVLSSGTFM  
 RGLIHIGTQNFSGGRLGDAASLGLSEDLKRLGFPLGRLKTGTPARLLASSIDFSVMEEQPGDHNVCVHR  
 NEMFVPTLPQ  
 VSCHITHTTDQTKDLITKNLHRSALYGGRIEGVGPRYCPSIEDKIVKFADKDRHHIFIEPEGLNTQEVYV  
 35 NGLSTSMFPD  
 VQYDIIRSVSGLENAIITRPAYAIEYDYVHGNIIFPSLESKLIIEGLFLCGQINGTTGYEAAAQGLIAGV  
 NAVNKVLRHP  
 PFVPSRQESYIGVMLDDLTTQVLDEPYRMFTSRAEHRLLLRQDNAGMRLSHYGHSLGILLSSERYAMFQEQ  
 40 KACIEQEKER  
 LSKTFRKYGDTVPLTKVLCRPEVSYQQLLTEFPADVRDLGPVVGASLEMEIKYSGYISRQQLIRSMER  
 SENISIPEDI  
 DYHSISALSLEAREKLSKFTPTRTIGSAARISGISVADIQVLMVSLKKDAH - (R<sub>2</sub>)<sub>n</sub>-Y

## Polypeptides

50 The polypeptides of the invention include the polypeptide of Table 1 [SEQ ID NO:2] (in particular the mature polypeptide) as well as polypeptides and fragments, particularly those which have the biological activity of gidA, and also those which have at least 70% identity to the polypeptide of Table 1 [SEQ ID NO:2] or the relevant portion, preferably at least 80% identity to the polypeptide of Table 1 [SEQ ID NO:2], and more preferably at least 90% similarity (more preferably at least 90% identity) to the polypeptide of Table 1 [SEQ ID NO:2] and still more preferably at least 95% similarity (still more preferably at least 95% identity) to the polypeptide of Table 1 [SEQ ID NO:2] and also include

55 portions of such polypeptides with such portion of the polypeptide generally containing at least 30 amino acids and more preferably at least 50 amino acids.

The invention also includes polypeptides of the formula set forth in Table 1 (D) wherein, at the amino terminus, X is hydrogen, and at the carboxyl terminus, Y is hydrogen or a metal, R<sub>1</sub> and R<sub>2</sub> is any amino acid residue, and n is an

integer between 1 and 1000. Any stretch of amino acid residues denoted by either R group, where R is greater than 1, may be either a heteropolymer or a homopolymer, preferably a heteropolymer.

A fragment is a variant polypeptide having an amino acid sequence that entirely is the same as part but not all of the amino acid sequence of the aforementioned polypeptides. As with gidAl polypeptides fragments may be "free-standing," or comprised within a larger polypeptide of which they form a part or region, most preferably as a single continuous region, a single larger polypeptide.

Preferred fragments include, for example, truncation polypeptides having a portion of the amino acid sequence of Table 1 [SEQ ID NO:2], or of variants thereof, such as a continuous series of residues that includes the amino terminus, or a continuous series of residues that includes the carboxyl terminus. Degradation forms of the polypeptides of the invention in a host cell, particularly a *Chlamydia trachomatis*, are also preferred. Further preferred are fragments characterized by structural or functional attributes such as fragments that comprise alpha-helix and alpha-helix forming regions, beta-sheet and beta-sheet-forming regions, turn and turn-forming regions, coil and coil-forming regions, hydrophilic regions, hydrophobic regions, alpha amphipathic regions, beta amphipathic regions, flexible regions, surface-forming regions, substrate binding region, and high antigenic index regions.

Also preferred are biologically active fragments which are those fragments that mediate activities of gidAl, including those with a similar activity or an improved activity, or with a decreased undesirable activity. Also included are those fragments that are antigenic or immunogenic in an animal, especially in a human. Particularly preferred are fragments comprising receptors or domains of enzymes that confer a function essential for viability of *Chlamydia trachomatis* or the ability to initiate, or maintain cause disease in an individual, particularly a human.

Variants that are fragments of the polypeptides of the invention may be employed for producing the corresponding full-length polypeptide by peptide synthesis; therefore, these variants may be employed as intermediates for producing the full-length polypeptides of the invention.

### Polynucleotides

Another aspect of the invention relates to isolated polynucleotides, including the full length gene, that encode the gidAl polypeptide having the deduced amino acid sequence of Table 1 [SEQ ID NO:2] and polynucleotides closely related thereto and variants thereof.

Using the information provided herein, such as the polynucleotide sequence set out in Table 1 [SEQ ID NO:1], a polynucleotide of the invention encoding gidAl polypeptide may be obtained using standard cloning and screening methods, such as those for cloning and sequencing chromosomal DNA fragments from bacteria using *Chlamydia trachomatis* D/UW-3/CX cells as starting material, followed by obtaining a full length clone. For example, to obtain a polynucleotide sequence of the invention, such as the sequence given in Table 1 [SEQ ID NO:1], typically a library of clones of chromosomal DNA of *Chlamydia trachomatis* D/UW-3/CX in *E. coli* or some other suitable host is probed with a radiolabeled oligonucleotide, preferably a 17-mer or longer, derived from a partial sequence. Clones carrying DNA identical to that of the probe can then be distinguished using stringent conditions. By sequencing the individual clones thus identified with sequencing primers designed from the original sequence it is then possible to extend the sequence in both directions to determine the full gene sequence. Conveniently, such sequencing is performed using denatured double stranded DNA prepared from a plasmid clone. Suitable techniques are described by Maniatis, T., Fritsch, E.F. and Sambrook et al., *MOLECULAR CLONING, A LABORATORY MANUAL*, 2nd Ed.; Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (1989). (see in particular Screening By Hybridization 1.90 and Sequencing Denatured Double-Stranded DNA Templates 13.70). Illustrative of the invention, the polynucleotide set out in Table 1 [SEQ ID NO:1] was discovered in a DNA library derived from *Chlamydia trachomatis* D/UW-3/CX.

The DNA sequence set out in Table 1 [SEQ ID NO:1] contains an open reading frame encoding a protein having about the number of amino acid residues set forth in Table 1 [SEQ ID NO:2] with a deduced molecular weight that can be calculated using amino acid residue molecular weight values well known in the art. The polynucleotide of SEQ ID NO: 1, between nucleotide number 535 through number 2364 encodes the polypeptide of SEQ ID NO:2. The stop codon begins at nucleotide number 2365 of SEQ ID NO: 1.

The gidAl protein of the invention is structurally related to other proteins of the gidA family, as shown by the results of sequencing the DNA encoding gidAl of the strain of the invention. The protein exhibits greatest homology to *Pseudomonas putida* gidA protein among known proteins. The gidAl protein of Table 1 [SEQ ID NO:2] has about 53% identity over its entire length and about 71% similarity over its entire length with the amino acid sequence of *Pseudomonas putida* gidA polypeptide.

The invention provides a polynucleotide sequence identical over its entire length to the coding sequence in Table 1 [SEQ ID NO:1]. Also provided by the invention is the coding sequence for the mature polypeptide or a fragment thereof, by itself as well as the coding sequence for the mature polypeptide or a fragment in a reading frame with other coding sequence, such as those encoding a leader or secretory sequence, a pre-, or pro- or prepro- protein sequence. The polynucleotide may also contain non-coding sequences, including for example, but not limited to non-coding 5'

and 3' sequences, such as the transcribed, non-translated sequences, termination signals, ribosome binding sites, sequences that stabilize mRNA, introns, polyadenylation signals, and additional coding sequence which encode additional amino acids. For example, a marker sequence that facilitates purification of the fused polypeptide can be encoded. In certain embodiments of the invention, the marker sequence is a hexa-histidine peptide, as provided in the pQE vector (Qiagen, Inc.) and described in Gentz *et al.*, *Proc. Natl. Acad. Sci. USA* 86: 821-824 (1989), or an HA tag (Wilson *et al.*, *Cell* 37: 767 (1984). Polynucleotides of the invention also include, but are not limited to, polynucleotides comprising a structural gene and its naturally associated sequences that control gene expression.

A preferred embodiment of the invention is the polynucleotide of comprising nucleotide 535 to 2364 set forth in SEQ ID NO:1 of Table 1 which encodes the gidAl polypeptide.

The invention also includes polynucleotides of the formula set forth in Table 1 (C) wherein, at the 5' end of the molecule, X is hydrogen, and at the 3' end of the molecule, Y is hydrogen or a metal, R<sub>1</sub> and R<sub>2</sub> is any nucleic acid residue, and n is an integer between 1 and 1000. Any stretch of nucleic acid residues denoted by either R group, where R is greater than 1, may be either a heteropolymer or a homopolymer, preferably a heteropolymer.

The term "polynucleotide encoding a polypeptide" as used herein encompasses polynucleotides that include a sequence encoding a polypeptide of the invention, particularly a bacterial polypeptide and more particularly a polypeptide of the *Chlamydia trachomatis* gidAl having the amino acid sequence set out in Table 1 [SEQ ID NO:2]. The term also encompasses polynucleotides that include a single continuous region or discontinuous regions encoding the polypeptide (for example, interrupted by integrated phage or an insertion sequence or editing) together with additional regions, that also may contain coding and/or non-coding sequences.

The invention further relates to variants of the polynucleotides described herein that encode for variants of the polypeptide having the deduced amino acid sequence of Table 1 [SEQ ID NO:2]. Variants that are fragments of the polynucleotides of the invention may be used to synthesize full-length polynucleotides of the invention.

Further particularly preferred embodiments are polynucleotides encoding gidAl variants, that have the amino acid sequence of gidAl polypeptide of Table 1 [SEQ ID NO:2] in which several, a few, 5 to 10, 1 to 5, 1 to 3, 2, 1 or no amino acid residues are substituted, deleted or added, in any combination. Especially preferred among these are silent substitutions, additions and deletions, that do not alter the properties and activities of gidAl.

Further preferred embodiments of the invention are polynucleotides that are at least 70% identical over their entire length to a polynucleotide encoding gidAl polypeptide having the amino acid sequence set out in Table 1 [SEQ ID NO:2], and polynucleotides that are complementary to such polynucleotides. Alternatively, most highly preferred are polynucleotides that comprise a region that is at least 80% identical over its entire length to a polynucleotide encoding gidAl polypeptide of the strain and polynucleotides complementary thereto. In this regard, polynucleotides at least 90% identical over their entire length to the same are particularly preferred, and among these particularly preferred polynucleotides, those with at least 95% are especially preferred. Furthermore, those with at least 97% are highly preferred among those with at least 95%, and among these those with at least 98% and at least 99% are particularly highly preferred, with at least 99% being the more preferred.

Preferred embodiments are polynucleotides that encode polypeptides that retain substantially the same biological function or activity as the mature polypeptide encoded by the DNA of Table 1 [SEQ ID NO:1].

The invention further relates to polynucleotides that hybridize to the herein above-described sequences. In this regard, the invention especially relates to polynucleotides that hybridize under stringent conditions to the herein above-described polynucleotides. As herein used, the terms "stringent conditions" and "stringent hybridization conditions" mean hybridization will occur only if there is at least 95% and preferably at least 97% identity between the sequences. An example of stringent hybridization conditions is overnight incubation at 42°C in a solution comprising: 50% formamide, 5x SSC (150mM NaCl, 15mM trisodium citrate), 50 mM sodium phosphate (pH7.6), 5x Denhardt's solution, 10% dextran sulfate, and 20 micrograms/ml denatured, sheared salmon sperm DNA, followed by washing the hybridization support in 0.1x SSC at about 65°C. Hybridization and wash conditions are well known and exemplified in Sambrook, *et al.*, *Molecular Cloning: A Laboratory Manual*, Second Edition, Cold Spring Harbor, N.Y., (1989), particularly Chapter 11 therein.

The invention also provides a polynucleotide consisting essentially of a polynucleotide sequence obtainable by screening an appropriate library containing the complete gene for a polynucleotide sequence set forth in SEQ ID NO: 1 under stringent hybridization conditions with a probe having the sequence of said polynucleotide sequence set forth in SEQ ID NO: 1 or a fragment thereof; and isolating said DNA sequence. Fragments useful for obtaining such a polynucleotide include, for example, probes and primers described elsewhere herein.

As discussed additionally herein regarding polynucleotide assays of the invention, for instance, polynucleotides of the invention as discussed above, may be used as a hybridization probe for RNA, cDNA and genomic DNA to isolate full-length cDNAs and genomic clones encoding gidAl and to isolate cDNA and genomic clones of other genes that have a high sequence similarity to the gidAl gene. Such probes generally will comprise at least 15 bases. Preferably, such probes will have at least 30 bases and may have at least 50 bases. Particularly preferred probes will have at least 30 bases and will have 50 bases or less.

For example, the coding region of the *gidA1* gene may be isolated by screening using the DNA sequence provided in SEQ ID NO: 1 to synthesize an oligonucleotide probe. A labeled oligonucleotide having a sequence complementary to that of a gene of the invention is then used to screen a library of cDNA, genomic DNA or mRNA to determine which members of the library the probe hybridizes to.

The polynucleotides and polypeptides of the invention may be employed, for example, as research reagents and materials for discovery of treatments of and diagnostics for disease, particularly human disease, as further discussed herein relating to polynucleotide assays.

Polynucleotides of the invention that are oligonucleotides derived from the sequences of SEQ ID NOS: 1 and/or 2 may be used in the processes herein as described, but preferably for PCR, to determine whether or not the polynucleotides identified herein in whole or in part are transcribed in bacteria in infected tissue. It is recognized that such sequences will also have utility in diagnosis of the stage of infection and type of infection the pathogen has attained.

The invention also provides polynucleotides that may encode a polypeptide that is the mature protein plus additional amino or carboxyl-terminal amino acids, or amino acids interior to the mature polypeptide (when the mature form has more than one polypeptide chain, for instance). Such sequences may play a role in processing of a protein from precursor to a mature form, may allow protein transport, may lengthen or shorten protein half-life or may facilitate manipulation of a protein for assay or production, among other things. As generally is the case *in vivo*, the additional amino acids may be processed away from the mature protein by cellular enzymes.

A precursor protein, having the mature form of the polypeptide fused to one or more prosequences may be an inactive form of the polypeptide. When prosequences are removed such inactive precursors generally are activated. Some or all of the prosequences may be removed before activation. Generally, such precursors are called proproteins.

In sum, a polynucleotide of the invention may encode a mature protein, a mature protein plus a leader sequence (which may be referred to as a preprotein), a precursor of a mature protein having one or more prosequences that are not the leader sequences of a preprotein, or a preproprotein, which is a precursor to a proprotein, having a leader sequence and one or more prosequences, which generally are removed during processing steps that produce active and mature forms of the polypeptide.

#### Vectors, host cells, expression

The invention also relates to vectors that comprise a polynucleotide or polynucleotides of the invention, host cells that are genetically engineered with vectors of the invention and the production of polypeptides of the invention by recombinant techniques. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the invention.

For recombinant production, host cells can be genetically engineered to incorporate expression systems or portions thereof or polynucleotides of the invention. Introduction of a polynucleotide into the host cell can be effected by methods described in many standard laboratory manuals, such as Davis et al., *BASIC METHODS IN MOLECULAR BIOLOGY*, (1986) and Sambrook et al., *MOLECULAR CLONING: A LABORATORY MANUAL*, 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989), such as, calcium phosphate transfection, DEAE-dextran mediated transfection, transfection, microinjection, cationic lipid-mediated transfection, electroporation, transduction, scrape loading, ballistic introduction and infection.

Representative examples of appropriate hosts include bacterial cells, such as streptococci, staphylococci, enterococci *E. coli*, streptomyces and *Bacillus subtilis* cells; fungal cells, such as yeast cells and *Aspergillus* cells; insect cells such as *Drosophila* S2 and *Spodoptera* Sf9 cells; animal cells such as CHO, COS, HeLa, C127, 3T3, BHK, 293 and Bowes melanoma cells; and plant cells.

A great variety of expression systems can be used to produce the polypeptides of the invention. Such vectors include, among others, chromosomal, episomal and virus-derived vectors, *e.g.*, vectors derived from bacterial plasmids, from bacteriophage, from transposons, from yeast episomes, from insertion elements, from yeast chromosomal elements, from viruses such as baculoviruses, papova viruses, such as SV40, vaccinia viruses, adenoviruses, fowl pox viruses, pseudorabies viruses and retroviruses, and vectors derived from combinations thereof, such as those derived from plasmid and bacteriophage genetic elements, such as cosmids and phagemids. The expression system constructs may contain control regions that regulate as well as engender expression. Generally, any system or vector suitable to maintain, propagate or express polynucleotides and/or to express a polypeptide in a host may be used for expression in this regard. The appropriate DNA sequence may be inserted into the expression system by any of a variety of well-known and routine techniques, such as, for example, those set forth in Sambrook et al., *MOLECULAR CLONING, A LABORATORY MANUAL*, (*supra*).

For secretion of the translated protein into the lumen of the endoplasmic reticulum, into the periplasmic space or into the extracellular environment, appropriate secretion signals may be incorporated into the expressed polypeptide. These signals may be endogenous to the polypeptide or they may be heterologous signals.

Polypeptides of the invention can be recovered and purified from recombinant cell cultures by well-known methods

including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography, and lectin chromatography. Most preferably, high performance liquid chromatography is employed for purification. Well known techniques for refolding protein may be employed to regenerate active conformation when the polypeptide is denatured during isolation and or purification.

### Diagnostic Assays

This invention is also related to the use of the gidAl polynucleotides of the invention for use as diagnostic reagents. Detection of gidAl in a eukaryote, particularly a mammal, and especially a human, will provide a diagnostic method for diagnosis of a disease. Eukaryotes (herein also "individual(s)"), particularly mammals, and especially humans, infected with an organism comprising the gidAl gene may be detected at the nucleic acid level by a variety of techniques.

Nucleic acids for diagnosis may be obtained from an infected individual's cells and tissues, such as bone, blood, muscle, cartilage, and skin. Genomic DNA may be used directly for detection or may be amplified enzymatically by using PCR or other amplification technique prior to analysis. RNA or cDNA may also be used in the same ways. Using amplification, characterization of the species and strain of prokaryote present in an individual, may be made by an analysis of the genotype of the prokaryote gene. Deletions and insertions can be detected by a change in size of the amplified product in comparison to the genotype of a reference sequence. Point mutations can be identified by hybridizing amplified DNA to labeled gidAl polynucleotide sequences. Perfectly matched sequences can be distinguished from mismatched duplexes by RNase digestion or by differences in melting temperatures. DNA sequence differences may also be detected by alterations in the electrophoretic mobility of the DNA fragments in gels, with or without denaturing agents, or by direct DNA sequencing. See, e.g., Myers et al., *Science*, 230: 1242 (1985). Sequence changes at specific locations also may be revealed by nuclease protection assays, such as RNase and S1 protection or a chemical cleavage method. See, e.g., Cotton et al., *Proc. Natl. Acad. Sci., USA*, 85: 4397-4401 (1988).

Cells carrying mutations or polymorphisms in the gene of the invention may also be detected at the DNA level by a variety of techniques, to allow for serotyping, for example. For example, RT-PCR can be used to detect mutations. It is particularly preferred to use RT-PCR in conjunction with automated detection systems, such as, for example, GeneScan. RNA or cDNA may also be used for the same purpose, PCR or RT-PCR. As an example, PCR primers complementary to a nucleic acid encoding gidAl can be used to identify and analyze mutations. These primers may be used for, among other things, amplifying gidAl DNA isolated from a sample derived from an individual. The primers may be used to amplify the gene isolated from an infected individual such that the gene may then be subject to various techniques for elucidation of the DNA sequence. In this way, mutations in the DNA sequence may be detected and used to diagnose infection and to serotype and/or classify the infectious agent.

The invention further provides a process for diagnosing, disease, preferably bacterial infections, more preferably infections by *Chlamydia trachomatis*, and most preferably classic ocular trachoma, inclusion conjunctivitis, genital trachoma, infant pneumonitis, Lymphogranuloma Venerium, incipient trachoma, keratitis, papillary hypertrophy, corneal infiltration, vulvovaginitis, ear infection, mucopurulent rhinitis, salpingitis, cervicitis, cervical follicles, prostatitis, proctitis, urethritis, lymphogranule inguinale, climatic bubo, tropical bubo, and esthiomene., comprising determining from a sample derived from an individual a increased level of expression of polynucleotide having the sequence of Table 1 [SEQ ID NO: 1]. Increased or decreased expression of gidAl polynucleotide can be measured using any one of the methods well known in the art for the quantitation of polynucleotides, such as, for example, amplification, PCR, RT-PCR, RNase protection, Northern blotting and other hybridization methods.

In addition, a diagnostic assay in accordance with the invention for detecting over-expression of gidA 1 protein compared to normal control tissue samples may be used to detect the presence of an infection, for example. Assay techniques that can be used to determine levels of a gidAl protein, in a sample derived from a host are well-known to those of skill in the art. Such assay methods include radioimmunoassays, competitive-binding assays, Western Blot analysis and ELISA assays.

### Antibodies

The polypeptides of the invention or variants thereof, or cells expressing them can be used as an immunogen to produce antibodies immunospecific for such polypeptides. "Antibodies" as used herein includes monoclonal and polyclonal antibodies, chimeric, single chain, simianized antibodies and humanized antibodies, as well as Fab fragments, including the products of an Fab immunoglobulin expression library.

Antibodies generated against the polypeptides of the invention can be obtained by administering the polypeptides or epitope-bearing fragments, analogues or cells to an animal, preferably a nonhuman, using routine protocols. For preparation of monoclonal antibodies, any technique known in the art that provides antibodies produced by continuous

cell line cultures can be used. Examples include various techniques, such as those in Kohler, G. and Milstein, C., *Nature* 256: 495-497 (1975); Kozbor *et al.*, *Immunology Today* 4: 72 (1983); Cole *et al.*, pg. 77-96 in *MONOCLONAL ANTIBODIES AND CANCER THERAPY*, Alan R. Liss, Inc. (1985).

Techniques for the production of single chain antibodies (U.S. Patent No. 4,946,778) can be adapted to produce single chain antibodies to polypeptides of this invention. Also, transgenic mice, or other organisms such as other mammals, may be used to express humanized antibodies.

Alternatively phage display technology may be utilized to select antibody genes with binding activities towards the polypeptide either from repertoires of PCR amplified v-genes of lymphocytes from humans screened for possessing anti-gidAI or from naive libraries (McCafferty, J. *et al.*, (1990), *Nature* 348, 552-554; Marks, J. *et al.*, (1992) *Biotechnology* 10, 779-783). The affinity of these antibodies can also be improved by chain shuffling (Clackson, T. *et al.*, (1991) *Nature* 352, 624-628).

If two antigen binding domains are present each domain may be directed against a different epitope - termed 'bispecific' antibodies.

The above-described antibodies may be employed to isolate or to identify clones expressing the polypeptides to purify the polypeptides by affinity chromatography.

Thus, among others, antibodies against gidAI- polypeptide may be employed to treat infections, particularly bacterial infections and especially classic ocular trachoma, inclusion conjunctivitis, genital trachoma, infant pneumonitis, Lymphogranuloma Venerium, incipient trachoma, keratitis, papillary hypertrophy, corneal infiltration, vulvovaginitis, ear infection, mucopurulent rhinitis, salpingitis, cervicitis, cervical follicles, prostatitis, proctitis, urethritis, lymphogranule inguinale, climatic bubo, tropical bubo, and esthiomene.

Polypeptide variants include antigenically, epitopically or immunologically equivalent variants that form a particular aspect of this invention. The term "antigenically equivalent derivative" as used herein encompasses a polypeptide or its equivalent which will be specifically recognized by certain antibodies which, when raised to the protein or polypeptide according to the invention, interfere with the immediate physical interaction between pathogen and mammalian host. The term "immunologically equivalent derivative" as used herein encompasses a peptide or its equivalent which when used in a suitable formulation to raise antibodies in a vertebrate, the antibodies act to interfere with the immediate physical interaction between pathogen and mammalian host.

The polypeptide, such as an antigenically or immunologically equivalent derivative or a fusion protein thereof is used as an antigen to immunize a mouse or other animal such as a rat or chicken. The fusion protein may provide stability to the polypeptide. The antigen may be associated, for example by conjugation, with an immunogenic carrier protein for example bovine serum albumin (BSA) or keyhole limpet haemocyanin (KLH). Alternatively a multiple antigenic peptide comprising multiple copies of the protein or polypeptide, or an antigenically or immunologically equivalent polypeptide thereof may be sufficiently antigenic to improve immunogenicity so as to obviate the use of a carrier.

Preferably, the antibody or variant thereof is modified to make it less immunogenic in the individual. For example, if the individual is human the antibody may most preferably be "humanized"; where the complementarity determining region(s) of the hybridoma-derived antibody has been transplanted into a human monoclonal antibody, for example as described in Jones, P. *et al.* (1986), *Nature* 321, 522-525 or Tempest *et al.*, (1991) *Biotechnology* 9, 266-273.

The use of a polynucleotide of the invention in genetic immunization will preferably employ a suitable delivery method such as direct injection of plasmid DNA into muscles (Wolff *et al.*, *Hum Mol Genet* 1992, 1:363, Manthorpe *et al.*, *Hum. Gene Ther.* 1993:4, 419), delivery of DNA complexed with specific protein carriers (Wu *et al.*, *J Biol Chem.* 1989: 264,16985), coprecipitation of DNA with calcium phosphate (Benvenisty & Reshef, *PNAS USA*, 1986:83,9551), encapsulation of DNA in various forms of liposomes (Kaneda *et al.*, *Science* 1989:243,375), particle bombardment (Tang *et al.*, *Nature* 1992, 356:152, Eisenbraun *et al.*, *DNA Cell Biol* 1993, 12:791) and *in vivo* infection using cloned retroviral vectors (Seeger *et al.*, *PNAS USA* 1984:81,5849).

#### Antagonists and agonists - assays and molecules

Polypeptides of the invention may also be used to assess the binding of small molecule substrates and ligands in, for example, cells, cell-free preparations, chemical libraries, and natural product mixtures. These substrates and ligands may be natural substrates and ligands or may be structural or functional mimetics. See, *e.g.*, Coligan *et al.*, *Current Protocols in Immunology* 1(2): Chapter 5 (1991).

The invention also provides a method of screening compounds to identify those which enhance (agonist) or block (antagonist) the action of gidAI polypeptides or polynucleotides, particularly those compounds that are bacteriostatic and/or bacteriocidal. The method of screening may involve high-throughput techniques. For example, to screen for agonists or antagonists, a synthetic reaction mix, a cellular compartment, such as a membrane, cell envelope or cell wall, or a preparation of any thereof, comprising gidAI polypeptide and a labeled substrate or ligand of such polypeptide is incubated in the absence or the presence of a candidate molecule that may be a gidAI agonist or antagonist. The ability of the candidate molecule to agonize or antagonize the gidAI polypeptide is reflected in decreased binding of

the labeled ligand or decreased production of product from such substrate. Molecules that bind gratuitously, i.e., without inducing the effects of gidAI polypeptide are most likely to be good antagonists. Molecules that bind well and increase the rate of product production from substrate are agonists. Detection of the rate or level of production of product from substrate may be enhanced by using a reporter system. Reporter systems that may be useful in this regard include but are not limited to colorimetric labeled substrate converted into product, a reporter gene that is responsive to changes in gidAI polynucleotide or polypeptide activity, and binding assays known in the art.

Another example of an assay for gidAI antagonists is a competitive assay that combines gidAI and a potential antagonist with gidAI-binding molecules, recombinant gidAI binding molecules, natural substrates or ligands, or substrate or ligand mimetics, under appropriate conditions for a competitive inhibition assay. The gidAI molecule can be labeled, such as by radioactivity or a colorimetric compound, such that the number of gidAI molecules bound to a binding molecule or converted to product can be determined accurately to assess the effectiveness of the potential antagonist.

Potential antagonists include small organic molecules, peptides, polypeptides and antibodies that bind to a polynucleotide or polypeptide of the invention and thereby inhibit or extinguish its activity. Potential antagonists also may be small organic molecules, a peptide, a polypeptide such as a closely related protein or antibody that binds the same sites on a binding molecule, such as a binding molecule, without inducing gidAI-induced activities, thereby preventing the action of gidAI by excluding gidAI from binding.

Potential antagonists include a small molecule that binds to and occupies the binding site of the polypeptide thereby preventing binding to cellular binding molecules, such that normal biological activity is prevented. Examples of small molecules include but are not limited to small organic molecules, peptides or peptide-like molecules. Other potential antagonists include antisense molecules (see Okano, *J. Neurochem.* 56: 560 (1991); *OLIGODEOXYNUCLEOTIDES AS ANTISENSE INHIBITORS OF GENE EXPRESSION*, CRC Press, Boca Raton, FL (1988), for a description of these molecules). Preferred potential antagonists include compounds related to and variants of gidAI.

Each of the DNA sequences provided herein may be used in the discovery and development of antibacterial compounds. The encoded protein, upon expression, can be used as a target for the screening of antibacterial drugs. Additionally, the DNA sequences encoding the amino terminal regions of the encoded protein or Shine-Delgarno or other translation facilitating sequences of the respective mRNA can be used to construct antisense sequences to control the expression of the coding sequence of interest.

The invention also provides the use of the polypeptide, polynucleotide or inhibitor of the invention to interfere with the initial physical interaction between a pathogen and mammalian host responsible for sequelae of infection. In particular the molecules of the invention may be used: in the prevention of adhesion of bacteria, in particular gram positive bacteria, to mammalian extracellular matrix proteins on in-dwelling devices or to extracellular matrix proteins in wounds; to block gidAI protein-mediated mammalian cell invasion by, for example, initiating phosphorylation of mammalian tyrosine kinases (Rosenshine *et al.*, *Infect. Immun.* 60:2211 (1992); to block bacterial adhesion between mammalian extracellular matrix proteins and bacterial gidAI proteins that mediate tissue damage and; to block the normal progression of pathogenesis in infections initiated other than by the implantation of in-dwelling devices or by other surgical techniques.

The antagonists and agonists of the invention may be employed, for instance, to inhibit and treat classic ocular trachoma, inclusion conjunctivitis, genital trachoma, infant pneumonitis, Lymphogranuloma Venerium, incipient trachoma, keratitis, papillary hypertrophy, corneal infiltration, vulvovaginitis, ear infection, mucopurulent rhinitis, salpingitis, cervicitis, cervical follicles, prostatitis, proctitis, urethritis, lymphogranule inguinale, climatic bubo, tropical bubo, and esthiomene.

## Vaccines

Another aspect of the invention relates to a method for inducing an immunological response in an individual, particularly a mammal which comprises inoculating the individual with gidAI, or a fragment or variant thereof, adequate to produce antibody and/or T cell immune response to protect said individual from infection, particularly bacterial infection and most particularly *Chlamydia trachomatis* infection. Also provided are methods whereby such immunological response slows bacterial replication. Yet another aspect of the invention relates to a method of inducing immunological response in an individual which comprises delivering to such individual a nucleic acid vector to direct expression of gidAI, or a fragment or a variant thereof, for expressing gidAI, or a fragment or a variant thereof *in vivo* in order to induce an immunological response, such as, to produce antibody, and/or T cell immune response, including, for example, cytokine-producing T cells or cytotoxic T cells, to protect said individual from disease, whether that disease is already established within the individual or not. One way of administering the gene is by accelerating it into the desired cells as a coating on particles or otherwise.

Such nucleic acid vector may comprise DNA, RNA, a modified nucleic acid, or a DNA/RNA hybrid.

A further aspect of the invention relates to an immunological composition which, when introduced into an individual

capable or having induced within it an immunological response, induces an immunological response in such individual to a gidAl or protein coded therefrom, wherein the composition comprises a recombinant gidAl or protein coded therefrom comprising DNA which codes for and expresses an antigen of said gidAl or protein coded therefrom. The immunological response may be used therapeutically or prophylactically and may take the form of antibody immunity or cellular immunity such as that arising from CTL or CD4+ T cells.

A gidAl polypeptide or a fragment thereof may be fused with co-protein which may not by itself produce antibodies, but is capable of stabilizing the first protein and producing a fused protein which will have immunogenic and protective properties. Thus fused recombinant protein, preferably further comprises an antigenic co-protein, such as lipoprotein D from *Hemophilus influenzae*, Glutathione-S-transferase (GST) or beta-galactosidase, relatively large co-proteins which solubilize the protein and facilitate production and purification thereof. Moreover, the co-protein may act as an adjuvant in the sense of providing a generalized stimulation of the immune system. The co-protein may be attached to either the amino or carboxy terminus of the first protein.

Provided by this invention are compositions, particularly vaccine compositions, and methods comprising the polypeptides or polynucleotides of the invention and immunostimulatory DNA sequences, such as those described in Sato, Y. *et al. Science* 273: 352 (1996).

Also, provided by this invention are methods using the described polynucleotide or particular fragments thereof which have been shown to encode non-variable regions of bacterial cell surface proteins in DNA constructs used in such genetic immunization experiments in animal models of infection with *Chlamydia trachomatis* will be particularly useful for identifying protein epitopes able to provoke a prophylactic or therapeutic immune response. It is believed that this approach will allow for the subsequent preparation of monoclonal antibodies of particular value from the requisite organ of the animal successfully resisting or clearing infection for the development of prophylactic agents or therapeutic treatments of bacterial infection, particularly *Chlamydia trachomatis* infection, in mammals, particularly humans.

The polypeptide may be used as an antigen for vaccination of a host to produce specific antibodies which protect against invasion of bacteria, for example by blocking adherence of bacteria to damaged tissue. Examples of tissue damage include wounds in skin or connective tissue caused, e.g., by mechanical, chemical or thermal damage or by implantation of indwelling devices, or wounds in the mucous membranes, such as the mouth, mammary glands, urethra or vagina.

The invention also includes a vaccine formulation which comprises an immunogenic recombinant protein of the invention together with a suitable carrier. Since the protein may be broken down in the stomach, it is preferably administered parenterally, including, for example, administration that is subcutaneous, intramuscular, intravenous, or intradermal. Formulations suitable for parenteral administration include aqueous and non-aqueous sterile injection solutions which may contain anti-oxidants, buffers, bacteriostats and solutes which render the formulation isotonic with the bodily fluid, preferably the blood, of the individual; and aqueous and non-aqueous sterile suspensions which may include suspending agents or thickening agents. The formulations may be presented in unit-dose or multi-dose containers, for example, sealed ampules and vials and may be stored in a freeze-dried condition requiring only the addition of the sterile liquid carrier immediately prior to use. The vaccine formulation may also include adjuvant systems for enhancing the immunogenicity of the formulation, such as oil-in water systems and other systems known in the art. The dosage will depend on the specific activity of the vaccine and can be readily determined by routine experimentation.

While the invention has been described with reference to certain gidAl protein, it is to be understood that this covers fragments of the naturally occurring protein and similar proteins with additions, deletions or substitutions which do not substantially affect the immunogenic properties of the recombinant protein.

#### Compositions, kits and administration

The invention also relates to compositions comprising the polynucleotide or the polypeptides discussed above or their agonists or antagonists. The polypeptides of the invention may be employed in combination with a non-sterile or sterile carrier or carriers for use with cells, tissues or organisms, such as a pharmaceutical carrier suitable for administration to a subject. Such compositions comprise, for instance, a media additive or a therapeutically effective amount of a polypeptide of the invention and a pharmaceutically acceptable carrier or excipient. Such carriers may include, but are not limited to, saline, buffered saline, dextrose, water, glycerol, ethanol and combinations thereof. The formulation should suit the mode of administration. The invention further relates to diagnostic and pharmaceutical packs and kits comprising one or more containers filled with one or more of the ingredients of the aforementioned compositions of the invention.

Polypeptides and other compounds of the invention may be employed alone or in conjunction with other compounds, such as therapeutic compounds.

The pharmaceutical compositions may be administered in any effective, convenient manner including, for instance, administration by topical, oral, anal, vaginal, intravenous, intraperitoneal, intramuscular, subcutaneous, intranasal or



intradermal routes among others.

In therapy or as a prophylactic, the active agent may be administered to an individual as an injectable composition, for example as a sterile aqueous dispersion, preferably isotonic.

Alternatively the composition may be formulated for topical application for example in the form of ointments, creams, lotions, eye ointments, eye drops, ear drops, mouthwash, impregnated dressings and sutures and aerosols, and may contain appropriate conventional additives, including, for example, preservatives, solvents to assist drug penetration, and emollients in ointments and creams. Such topical formulations may also contain compatible conventional carriers, for example cream or ointment bases, and ethanol or oleyl alcohol for lotions. Such carriers may constitute from about 1% to about 98% by weight of the formulation; more usually they will constitute up to about 80% by weight of the formulation.

For administration to mammals, and particularly humans, it is expected that the daily dosage level of the active agent will be from 0.01 mg/kg to 10 mg/kg, typically around 1 mg/kg. The physician in any event will determine the actual dosage which will be most suitable for an individual and will vary with the age, weight and response of the particular individual. The above dosages are exemplary of the average case. There can, of course, be individual instances where higher or lower dosage ranges are merited, and such are within the scope of this invention.

In-dwelling devices include surgical implants, prosthetic devices and catheters, i.e., devices that are introduced to the body of an individual and remain in position for an extended time. Such devices include, for example, artificial joints, heart valves, pacemakers, vascular grafts, vascular catheters, cerebrospinal fluid shunts, urinary catheters, continuous ambulatory peritoneal dialysis (CAPD) catheters.

The composition of the invention may be administered by injection to achieve a systemic effect against relevant bacteria shortly before insertion of an in-dwelling device. Treatment may be continued after surgery during the in-body time of the device. In addition, the composition could also be used to broaden perioperative cover for any surgical technique to prevent bacterial wound infections, especially *Chlamydia trachomatis* wound infections.

Many orthopaedic surgeons consider that humans with prosthetic joints should be considered for antibiotic prophylaxis before dental treatment that could produce a bacteremia. Late deep infection is a serious complication sometimes leading to loss of the prosthetic joint and is accompanied by significant morbidity and mortality. It may therefore be possible to extend the use of the active agent as a replacement for prophylactic antibiotics in this situation.

In addition to the therapy described above, the compositions of this invention may be used generally as a wound treatment agent to prevent adhesion of bacteria to matrix proteins exposed in wound tissue and for prophylactic use in dental treatment as an alternative to, or in conjunction with, antibiotic prophylaxis.

Alternatively, the composition of the invention may be used to bathe an indwelling device immediately before insertion. The active agent will preferably be present at a concentration of 1µg/ml to 10mg/ml for bathing of wounds or indwelling devices.

A vaccine composition is conveniently in injectable form. Conventional adjuvants may be employed to enhance the immune response. A suitable unit dose for vaccination is 0.5-5 microgram/kg of antigen, and such dose is preferably administered 1-3 times and with an interval of 1-3 weeks. With the indicated dose range, no adverse toxicological effects will be observed with the compounds of the invention which would preclude their administration to suitable individuals.

Each reference disclosed herein is incorporated by reference herein in its entirety. Any patent application to which this application claims priority is also incorporated by reference herein in its entirety.

## EXAMPLES

The examples below are carried out using standard techniques, which are well known and routine to those of skill in the art, except where otherwise described in detail. The examples are illustrative, but do not limit the invention.

### Example 1 Strain selection, Library Production and Sequencing

The polynucleotide having the DNA sequence given in SEQ ID NO: 1 is obtained, for example from a library of clones of chromosomal DNA of *Chlamydia trachomatis* in *E. coli*. The sequencing data from two or more clones containing overlapping *Chlamydia trachomatis* DNAs is used to construct the contiguous DNA sequence in SEQ ID NO: 1. Libraries may be prepared by routine methods, for example:

#### Methods 1, 2 and 3 below.

Total cellular DNA is isolated from *Chlamydia trachomatis* D/UW-3/CX according to standard procedures and size-fractionated by either of two methods.

**Method 1**

Total cellular DNA is mechanically sheared by passage through a needle in order to size-fractionate according to standard procedures. DNA fragments of up to 11kbp in size are rendered blunt by treatment with exonuclease and DNA polymerase, and EcoRI linkers added. Fragments are ligated into the vector Lambda ZapII that has been cut with EcoRI, the library packaged by standard procedures and *E.coli* infected with the packaged library. The library is amplified by standard procedures.

**Method 2**

Total cellular DNA is partially hydrolyzed with a one or a combination of restriction enzymes appropriate to generate a series of fragments for cloning into library vectors (e.g., RsaI, PstI, AluI, Bsh1235I), and such fragments are size-fractionated according to standard procedures. EcoRI linkers are ligated to the DNA and the fragments then ligated into the vector Lambda ZapII that have been cut with EcoRI, the library packaged by standard procedures, and *E.coli* infected with the packaged library. The library is amplified by standard procedures.

**Method 3**

Total cellular DNA is mechanically or enzymatically fragmented to size-fractionate according to standard procedures. DNA fragments of about 1kbp in size, after preparing their ends using standard procedures, are ligated into M13 vector using standard procedures. M13 is introduced into *E.coli* host, such as NM522 (available commercially). Clones with inserts are sequenced using standard procedures.

Annex to the description

SEQUENCE LISTING

(1) GENERAL INFORMATION

(i) APPLICANT: SmithKline Beecham Corporation & SmithKline Beecham Plc

(ii) TITLE OF THE INVENTION: NOVEL GLYCOGEN PHOSPHORYL  
ASE

(iii) NUMBER OF SEQUENCES: 6

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: SmithKline Beecham, Corporate Intellectual Property

(B) STREET: Two New Horizons Court

(C) CITY: Brentford

(D) STATE: Middlesex

(E) COUNTRY: United Kingdom

(F) ZIP: TW8 9EP

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Diskette

(B) COMPUTER: IBM Compatible

(C) OPERATING SYSTEM: DOS

(D) SOFTWARE: FastSEQ for Windows Version 2.0

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER:

(B) FILING DATE:

(C) CLASSIFICATION:

## (vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER:

(B) FILING DATE:

## (viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: CONNELL, Anthony Christopher

(B) REGISTRATION NUMBER: 5630 &amp; 26758

(C) REFERENCE/DOCKET NUMBER: GM10043

## (ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: +44 1279 644 395

(B) TELEFAX: +44 181 975 6294

(C) TELEX:

## (2) INFORMATION FOR SEQ ID NO:1:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2259 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

ATGTTATCAC TACAAGAATT TGTACAAAAT CGTTACAATA AAACCATTGC  
AGAATGTAGC 60  
AATGAAGAGC TTTACCTTGC TCTTCTTAAC TACAGCAAGC TTGCAAGCAG  
CCAAAAACCA 120

GTCAACACTG GTAAGAAAAA AGTTTACTAC ATCTCAGCTG AGTTCTTGAT  
TGGTAAACTC 180

5 TTGTCAAACA ACTTGATTAA CCTTGGTCTT TACGACGATG TTA AAAAAGA  
ACTTGCAGCT 240

GCAGGTAAAG ACTTGATCGA AGTTGAAGAA GTTGAATTGG AACCATCTCT  
10 TGGTAATGGT 300

GGTTTGGGAC GTTTGGCTGC CTGCTTTATC GACTCAATTG CTACTCTTGG  
TTTGAATGGT 360

15 GACGGTGTTG GTCTTAACTA CCACTTTGGT CTTTCCAAC AAGTTCTTAA  
AAACAACCAA 420

CAAGAAACAA TTCCAAATGC ATGGTTGACA GAGCAAAACT GGTGTTGTCG  
20 CTCAAGCCGT 480

AGCTACCAAG TACCATTGTC AGACTTTACT TTGACATCAA CTCTTTACGA  
TATTGATGTT 540

25 ACTGGTTATG AAACAGCGAC TAAAAACCGC TTGCGTTTGT TTGACTTGGA  
TTCAGTTGAT 600

TCTTCTATTA TAAAGATGG TATCAACTTT GACAAGACAG ATATCGCTCG  
30 CAACTTGACT 660

CTCTTCCTTT ACCCAGATGA TAGTGACCGT CAAGGTGAAT TGCTCCGTAT  
CTTCCAACAA 720

35 TACTTCATGG TTTCAAACGG TGCGCAATTG ATCATCGACG AAGCAATCGA  
AAAAGGAAGC 780

AACTTGCATG ACCTTGCTGA CTACGCAGTT GTCCAAATCA ACGATACTCA  
40 CCCATCAATG 840

GTGATTCCTG AATTGATTCG TCTTTTGACT GCACGTGGTA TCGAGCTTGA  
CGAAGCAATC 900

45 TCAATTGTTC GTAGCATGAC TGCCTACACT AACCACACAA TCCTTGCTGA  
GGCGCTTGAA 960

AAATGGCCTC TTGAATTCTT GCAAGAAGTG GTTCCTCACT TGGTACCAAT  
50 CATCGAAGAA 1020

TTGGACCGTC GTGTGAAGGC AGAGTACAAA GATCCAGCTG TTCAAATCAT  
CGATGAGAGC 1080

GGACGTGTTT ACATGGCTCA CATGGATATC CACTACGGAT ACAGTGTTAA  
CGGGGTTGCA 1140

5 GCACTTCATA CTGAAATCTT GAAAAATTCT GAGTTGAAAG CCTTCTACGA  
CCTTTACCCA 1200

GAAAAGTTCA ACAACAAAAC AAACGGTATC ACTTTCCGTC GTTGGCTTAT  
10 GCATGCTAAC 1260

CCAAGATTGT CTCACTACTT GGATGAGATT CTTGGAGATG GTTGGCACCA  
TGAAGCAGAT 1320

15 GAGCTTGAAA AACTTTTGTC TTATGAAGAC AAAGCAGCTG TCAAAGAAAA  
ATTGGAAAGC 1380

ATCAAGGCTC ACAACAAACG TAAATTGGCT CGTCACTTGA AAGAACACCA  
20 AGGTGTGGAA 1440

ATCAATCCAA ATTCTATCTT TGATATCCAA ATCAAACGTC TTCACGAGTA  
CAAACGCCAA 1500

25 CAAATGAACG CTTTGTACGT GATCCACAAA TACCTTGACA TCAAAGCTGG  
TAACATCCCT 1560

GCTCGTCCAA TCACAATCTT CTTTGGTGGT AAAGCAGCTC CAGCCTACAC  
30 AATCGCTCAA 1620

GACATTATCC ATTTAATCCT TTGCATGTCA GAAGTTATTG CTAACGATCC  
AGCAGTAGCT 1680

35 CCACACTTGC AAGTAGTTAT GGTGAAAAC TACAACGTTA CTGCAGCAAG  
TTTCCTTATC 1740

CCAGCATGTG ATATCTCAGA ACAAATCTCA CTTGCTTCTA AAGAAGCTTC  
40 AGGTACTGGT 1800

AACATGAAAT TCATGTTGAA CGGAGCTTTG ACACTTGGTA CTATGGACGG  
TGCTAACGTG 1860

45 GAAATCGCTG AGTTGGTTGG AGAAGAAAAC ATCTACATCT TCGGTGAAGA  
TTCAGAAACT 1920

GTTATCGACC TTTACGCAAA AGCAGCTTAC AAATCAAGCG AATTCTACGC  
50 TCGTGAAGCT 1980

ATCAAACCAT TGGTTGACTT CATCGTTAGT GATGCAGTTC TTGCAGCTGG  
AAACAAAGAG 2040

CGCTTGGAAC GTCTTTACAA TGAATTGATC AACAAAGACT GGTTTCATGAC  
TCTTCTTGAC 2100  
5 TTGGAAGACT ACATCAAAGT CAAAGAGCAA ATGCTTGCTG ACTACGAAGA  
CCGTGACGCA 2160  
TGGTTGGATA AAGTCATCGT TAACATTTCT AAAGCAGGAT TCTTCTCATC  
10 TGACCGTACA 2220  
ATCGCTCAGT ATAACGAAGA CATCTGGCAC TTGAACTAA 2259

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 752 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Leu Ser Leu Gln Glu Phe Val Gln Asn Arg Tyr Asn Lys Thr Ile

1 5 10 15

Ala Glu Cys Ser Asn Glu Glu Leu Tyr Leu Ala Leu Leu Asn Tyr Ser

20 25 30

Lys Leu Ala Ser Ser Gln Lys Pro Val Asn Thr Gly Lys Lys Lys Val

35 40 45

Tyr Tyr Ile Ser Ala Glu Phe Leu Ile Gly Lys Leu Leu Ser Asn Asn

50 55 60

Leu Ile Asn Leu Gly Leu Tyr Asp Asp Val Lys Lys Glu Leu Ala Ala

65 70 75 80

Ala Gly Lys Asp Leu Ile Glu Val Glu Glu Val Glu Leu Glu Pro Ser

85 90 95

Leu Gly Asn Gly Gly Leu Gly Arg Leu Ala Ala Cys Phe Ile Asp Ser

100            105            110  
 Ile Ala Thr Leu Gly Leu Asn Gly Asp Gly Val Gly Leu Asn Tyr His  
 5            115            120            125  
 Phe Gly Leu Phe Gln Gln Val Leu Lys Asn Asn Gln Gln Glu Thr Ile  
 130            135            140  
 10 Pro Asn Ala Trp Leu Thr Glu Gln Asn Trp Leu Val Arg Ser Ser Arg  
 145            150            155            160  
 Ser Tyr Gln Val Pro Phe Ala Asp Phe Thr Leu Thr Ser Thr Leu Tyr  
 15            165            170            175  
 Asp Ile Asp Val Thr Gly Tyr Glu Thr Ala Thr Lys Asn Arg Leu Arg  
 180            185            190  
 20 Leu Phe Asp Leu Asp Ser Val Asp Ser Ser Ile Ile Lys Asp Gly Ile  
 195            200            205  
 Asn Phe Asp Lys Thr Asp Ile Ala Arg Asn Leu Thr Leu Phe Leu Tyr  
 25            210            215            220  
 Pro Asp Asp Ser Asp Arg Gln Gly Glu Leu Leu Arg Ile Phe Gln Gln  
 225            230            235            240  
 30 Tyr Phe Met Val Ser Asn Gly Ala Gln Leu Ile Ile Asp Glu Ala Ile  
 245            250            255  
 Glu Lys Gly Ser Asn Leu His Asp Leu Ala Asp Tyr Ala Val Val Gln  
 260            265            270  
 35 Ile Asn Asp Thr His Pro Ser Met Val Ile Pro Glu Leu Ile Arg Leu  
 275            280            285  
 40 Leu Thr Ala Arg Gly Ile Glu Leu Asp Glu Ala Ile Ser Ile Val Arg  
 290            295            300  
 Ser Met Thr Ala Tyr Thr Asn His Thr Ile Leu Ala Glu Ala Leu Glu  
 305            310            315            320  
 45 Lys Trp Pro Leu Glu Phe Leu Gln Glu Val Val Pro His Leu Val Pro  
 325            330            335  
 50 Ile Ile Glu Glu Leu Asp Arg Arg Val Lys Ala Glu Tyr Lys Asp Pro  
 340            345            350  
 Ala Val Gln Ile Ile Asp Glu Ser Gly Arg Val His Met Ala His Met



355            360            365  
 Asp Ile His Tyr Gly Tyr Ser Val Asn Gly Val Ala Ala Leu His Thr  
 5            370            375            380  
 Glu Ile Leu Lys Asn Ser Glu Leu Lys Ala Phe Tyr Asp Leu Tyr Pro  
 385            390            395            400  
 10            Glu Lys Phe Asn Asn Lys Thr Asn Gly Ile Thr Phe Arg Arg Trp Leu  
 405            410            415  
 Met His Ala Asn Pro Arg Leu Ser His Tyr Leu Asp Glu Ile Leu Gly  
 15            420            425            430  
 Asp Gly Trp His His Glu Ala Asp Glu Leu Glu Lys Leu Leu Ser Tyr  
 435            440            445  
 20            Glu Asp Lys Ala Ala Val Lys Glu Lys Leu Glu Ser Ile Lys Ala His  
 450            455            460  
 Asn Lys Arg Lys Leu Ala Arg His Leu Lys Glu His Gln Gly Val Glu  
 25            465            470            475            480  
 Ile Asn Pro Asn Ser Ile Phe Asp Ile Gln Ile Lys Arg Leu His Glu  
 485            490            495  
 30            Tyr Lys Arg Gln Gln Met Asn Ala Leu Tyr Val Ile His Lys Tyr Leu  
 500            505            510  
 Asp Ile Lys Ala Gly Asn Ile Pro Ala Arg Pro Ile Thr Ile Phe Phe  
 35            515            520            525  
 Gly Gly Lys Ala Ala Pro Ala Tyr Thr Ile Ala Gln Asp Ile Ile His  
 530            535            540  
 40            Leu Ile Leu Cys Met Ser Glu Val Ile Ala Asn Asp Pro Ala Val Ala  
 545            550            555            560  
 Pro His Leu Gln Val Val Met Val Glu Asn Tyr Asn Val Thr Ala Ala  
 45            565            570            575  
 Ser Phe Leu Ile Pro Ala Cys Asp Ile Ser Glu Gln Ile Ser Leu Ala  
 580            585            590  
 50            Ser Lys Glu Ala Ser Gly Thr Gly Asn Met Lys Phe Met Leu Asn Gly  
 595            600            605  
 Ala Leu Thr Leu Gly Thr Met Asp Gly Ala Asn Val Glu Ile Ala Glu  
 55

610 615 620  
Leu Val Gly Glu Glu Asn Ile Tyr Ile Phe Gly Glu Asp Ser Glu Thr  
5 625 630 635 640  
Val Ile Asp Leu Tyr Ala Lys Ala Ala Tyr Lys Ser Ser Glu Phe Tyr  
645 650 655  
10 Ala Arg Glu Ala Ile Lys Pro Leu Val Asp Phe Ile Val Ser Asp Ala  
660 665 670  
Val Leu Ala Ala Gly Asn Lys Glu Arg Leu Glu Arg Leu Tyr Asn Glu  
15 675 680 685  
Leu Ile Asn Lys Asp Trp Phe Met Thr Leu Leu Asp Leu Glu Asp Tyr  
690 695 700  
20 Ile Lys Val Lys Glu Gln Met Leu Ala Asp Tyr Glu Asp Arg Asp Ala  
705 710 715 720  
Trp Leu Asp Lys Val Ile Val Asn Ile Ser Lys Ala Gly Phe Phe Ser  
25 725 730 735  
Ser Asp Arg Thr Ile Ala Gln Tyr Asn Glu Asp Ile Trp His Leu Asn  
740 745 750  
30

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2256 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

ATGTTATCAC TACAAGAATT TGTACAAAAT CGTTACAATA AAACCATTGC  
AGAATGTAGC 60

AATGAAGAGC TTTACCTTGC TCTTCTTAAC TACAGCAAGC TTGCAAGCAG  
 CCAAAAACCA 120  
 5 GTCAACACTG GTAAGAAAAA AGTTTACTAC ATCTCAGCTG AGTTCTTGAT  
 TGGTAAACTC 180  
 TTGTCAAACA ACTTGATTAA CCTTGGTCTT TACGACGATG TAAAAAAGA  
 10 ACTTGCAGCT 240  
 GCAGGTAAAG ACTTGATCGA AGTTGAAGAA GTTGAATTGG AACCATCTCT  
 TGGTAATGGT 300  
 15 GGTTTGGGAC GTTTGGCTGC CTGCTTTATC GACTCAATTG CTACTCTTGG  
 TTTGAATGGT 360  
 GACGGTGTTG GTCTTAACTA CCACTTTGGT CTTTCCAAAC AAGTTCTTAA  
 20 AAACAACCAA 420  
 CAAGAAACAA TTCCAAATGC ATGGTTGACA GAGCAAAACT GGTTGGTTCG  
 CTCAAGCCGT 480  
 25 AGCTACCAAG TACCATTGTC AGACTTTACT TTGACATCAA CTCTTTACGA  
 TATTGATGTT 540  
 ACTGGTTATG AAACAGCGAC TAAAAACCGC TTGCGTTTGT TTGACTTGGA  
 30 TTCAGTTGAT 600  
 TCTTCTATTA TTAAGATGG TATCAACTTT GACAAGACAG ATATCGCTCG  
 CAACTTGACT 660  
 35 CTCTTCCTTT ACCCAGATGA TAGTGACCGT CAAGGTGAAT TGCTCCGTAT  
 CTTCCAACAA 720  
 TACTTCATGG TTTCAAACGG TGCGCAATTG ATCATCGACG AAGCAATCGA  
 40 AAAAGGAAGC 780  
 AACTTGCAATG ACCTTGCTGA CTACGCAGTT GTCCAAATCA ACGATACTCA  
 CCCATCAATG 840  
 45 GTGATTCCTG AATTGATTCTG TCTTTTGACT GCACGTGGTA TCGAGCTTGA  
 CGAAGCAATC 900  
 TCAATTGTTC GTAGCATGAC TGCCTACACT AACCACACAA TCCTTGCTGA  
 50 GGCGCTTGAA 960  
 AAATGGCCTC TTGAATTCTT GCAAGAAGTG GTTCCTCACT TGGTACCAAT  
 CATCGAAGAA 1020

TTGGACCGTC GTGTGAAGGC AGAGTACAAA GATCCAGCTG TTCAAATCAT  
CGATGAGAGC 1080

5 GGACGTGTTC ACATGGCTCA CATGGATATC CACTACGGAT ACAGTGTTAA  
CGGGGTTGCA 1140

10 GCACTTCATA CTGAAATCTT GAAAAATTCT GAGTTGAAAG CCTTCTACGA  
CCTTTACCCA 1200

GAAAAGTTCA ACAACAAAAC AAACGGTATC ACTTTCCGTC GTTGGCTTAT  
GCATGCTAAC 1260

15 CCAAGATTGT CTCACTACTT GGATGAGATT CTTGGAGATG GTTGGCACCA  
TGAAGCAGAT 1320

20 GAGCTTGAAA AACTTTTGTC TTATGAAGAC AAAGCAGCTG TCAAAGAAAA  
ATTGGAAAGC 1380

ATCAAGGCTC ACAACAAACG TAAATTGGCT CGTCACTTGA AAGAACACCA  
AGGTGTGGAA 1440

25 ATCAATCCAA ATTCTATCTT TGATATCCAA ATCAAACGTC TTCACGAGTA  
CAAACGCCAA 1500

30 CAAATGAACG CTTTGTACGT GATCCACAAA TACCTTGACA TCAAAGCTGG  
TAACATCCCT 1560

GCTCGTCCAA TCACAATCTT CTTTGGTGGT AAAGCAGCTC CAGCCTACAC  
AATCGCTCAA 1620

35 GACATTATCC ATTTAATCCT TTGCATGTCA GAAGTTATTG CTAACGATCC  
AGCAGTAGCT 1680

40 CCACACTTGC AAGTAGTTAT GGTTGAAAAC TACAACGTTA CTGCAGCAAG  
TTTCCTTATC 1740

CCAGCATGTG ATATCTCAGA ACAAATCTCA CTTGCTTCTA AAGAAGCTTC  
AGGTACTGGT 1800

45 AACATGAAAT TCATGTTGAA CGGAGCTTTG ACACTTGGTA CTATGGACGG  
TGCTAACGTG 1860

50 GAAATCGCTG AGTTGGTTGG AGAAGAAAAC ATCTACATCT TCGGTGAAGA  
TTCAGAAACT 1920

GTTATCGACC TTTACGCAA AGCAGCTTAC AAATCAAGCG AATTCTACGC  
TCGTGAAGCT 1980

ATCAAACCAT TGGTTGACTT CATCGTTAGT GATGCAGTTC TTGCAGCTGG  
AAACAAAGAG 2040

5 CGCTTGGAAC GTCTTTACAA TGAATTGATC AACAAAGACT GGTTCATGAC  
TCTTCTTGAC 2100

10 TTGGAAGACT ACATCAAAGT CAAAGAGCAA ATGCTTGCTG ACTACGAAGA  
CCGTGACGCA 2160

TGGTTGGATA AAGTCATCGT TAACATTTCT AAAGCAGGAT TCTTCTCATC  
TGACCGTACA 2220

15 ATCGCTCAGT ATAACGAAGA CATCTGGCAC TTGAAC 2256

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 752 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Met Leu Ser Leu Gln Glu Phe Val Gln Asn Arg Tyr Asn Lys Thr Ile

1 5 10 15

Ala Glu Cys Ser Asn Glu Glu Leu Tyr Leu Ala Leu Leu Asn Tyr Ser

20 25 30

Lys Leu Ala Ser Ser Gln Lys Pro Val Asn Thr Gly Lys Lys Lys Val

35 40 45

Tyr Tyr Ile Ser Ala Glu Phe Leu Ile Gly Lys Leu Leu Ser Asn Asn

50 55 60

Leu Ile Asn Leu Gly Leu Tyr Asp Asp Val Lys Lys Glu Leu Ala Ala

65 70 75 80

Ala Gly Lys Asp Leu Ile Glu Val Glu Glu Val Glu Leu Glu Pro Ser

85                      90                      95  
 Leu Gly Asn Gly Gly Leu Gly Arg Leu Ala Ala Cys Phe Ile Asp Ser  
 5                      100                      105                      110  
 Ile Ala Thr Leu Gly Leu Asn Gly Asp Gly Val Gly Leu Asn Tyr His  
 115                      120                      125  
 10 Phe Gly Leu Phe Gln Gln Val Leu Lys Asn Asn Gln Gln Glu Thr Ile  
 130                      135                      140  
 Pro Asn Ala Trp Leu Thr Glu Gln Asn Trp Leu Val Arg Ser Ser Arg  
 15 145                      150                      155                      160  
 Ser Tyr Gln Val Pro Phe Ala Asp Phe Thr Leu Thr Ser Thr Leu Tyr  
 165                      170                      175  
 20 Asp Ile Asp Val Thr Gly Tyr Glu Thr Ala Thr Lys Asn Arg Leu Arg  
 180                      185                      190  
 Leu Phe Asp Leu Asp Ser Val Asp Ser Ser Ile Ile Lys Asp Gly Ile  
 25 195                      200                      205  
 Asn Phe Asp Lys Thr Asp Ile Ala Arg Asn Leu Thr Leu Phe Leu Tyr  
 210                      215                      220  
 30 Pro Asp Asp Ser Asp Arg Gln Gly Glu Leu Leu Arg Ile Phe Gln Gln  
 225                      230                      235                      240  
 Tyr Phe Met Val Ser Asn Gly Ala Gln Leu Ile Ile Asp Glu Ala Ile  
 35 245                      250                      255  
 Glu Lys Gly Ser Asn Leu His Asp Leu Ala Asp Tyr Ala Val Val Gln  
 260                      265                      270  
 40 Ile Asn Asp Thr His Pro Ser Met Val Ile Pro Glu Leu Ile Arg Leu  
 275                      280                      285  
 Leu Thr Ala Arg Gly Ile Glu Leu Asp Glu Ala Ile Ser Ile Val Arg  
 290                      295                      300  
 45 Ser Met Thr Ala Tyr Thr Asn His Thr Ile Leu Ala Glu Ala Leu Glu  
 305                      310                      315                      320  
 50 Lys Trp Pro Leu Glu Phe Leu Gln Glu Val Val Pro His Leu Val Pro  
 325                      330                      335  
 Ile Ile Glu Glu Leu Asp Arg Arg Val Lys Ala Glu Tyr Lys Asp Pro  
 55

340            345            350  
 Ala Val Gln Ile Ile Asp Glu Ser Gly Arg Val His Met Ala His Met  
 5            355            360            365  
 Asp Ile His Tyr Gly Tyr Ser Val Asn Gly Val Ala Ala Leu His Thr  
 370            375            380  
 10            Glu Ile Leu Lys Asn Ser Glu Leu Lys Ala Phe Tyr Asp Leu Tyr Pro  
 385            390            395            400  
 Glu Lys Phe Asn Asn Lys Thr Asn Gly Ile Thr Phe Arg Arg Trp Leu  
 15            405            410            415  
 Met His Ala Asn Pro Arg Leu Ser His Tyr Leu Asp Glu Ile Leu Gly  
 420            425            430  
 20            Asp Gly Trp His His Glu Ala Asp Glu Leu Glu Lys Leu Leu Ser Tyr  
 435            440            445  
 Glu Asp Lys Ala Ala Val Lys Glu Lys Leu Glu Ser Ile Lys Ala His  
 25            450            455            460  
 Asn Lys Arg Lys Leu Ala Arg His Leu Lys Glu His Gln Gly Val Glu  
 465            470            475            480  
 30            Ile Asn Pro Asn Ser Ile Phe Asp Ile Gln Ile Lys Arg Leu His Glu  
 485            490            495  
 Tyr Lys Arg Gln Gln Met Asn Ala Leu Tyr Val Ile His Lys Tyr Leu  
 35            500            505            510  
 Asp Ile Lys Ala Gly Asn Ile Pro Ala Arg Pro Ile Thr Ile Phe Phe  
 515            520            525  
 40            Gly Gly Lys Ala Ala Pro Ala Tyr Thr Ile Ala Gln Asp Ile Ile His  
 530            535            540  
 Leu Ile Leu Cys Met Ser Glu Val Ile Ala Asn Asp Pro Ala Val Ala  
 45            545            550            555            560  
 Pro His Leu Gln Val Val Met Val Glu Asn Tyr Asn Val Thr Ala Ala  
 565            570            575  
 50            Ser Phe Leu Ile Pro Ala Cys Asp Ile Ser Glu Gln Ile Ser Leu Ala  
 580            585            590  
 Ser Lys Glu Ala Ser Gly Thr Gly Asn Met Lys Phe Met Leu Asn Gly  
 55

595                      600                      605

Ala Leu Thr Leu Gly Thr Met Asp Gly Ala Asn Val Glu Ile Ala Glu

610                      615                      620

Leu Val Gly Glu Glu Asn Ile Tyr Ile Phe Gly Glu Asp Ser Glu Thr

625                      630                      635                      640

Val Ile Asp Leu Tyr Ala Lys Ala Ala Tyr Lys Ser Ser Glu Phe Tyr

645                      650                      655

Ala Arg Glu Ala Ile Lys Pro Leu Val Asp Phe Ile Val Ser Asp Ala

660                      665                      670

Val Leu Ala Ala Gly Asn Lys Glu Arg Leu Glu Arg Leu Tyr Asn Glu

675                      680                      685

Leu Ile Asn Lys Asp Trp Phe Met Thr Leu Leu Asp Leu Glu Asp Tyr

690                      695                      700

Ile Lys Val Lys Glu Gln Met Leu Ala Asp Tyr Glu Asp Arg Asp Ala

705                      710                      715                      720

Trp Leu Asp Lys Val Ile Val Asn Ile Ser Lys Ala Gly Phe Phe Ser

725                      730                      735

Ser Asp Arg Thr Ile Ala Gln Tyr Asn Glu Asp Ile Trp His Leu Asn

740                      745                      750

## (2) INFORMATION FOR SEQ ID NO:5:

### (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 23 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

### (xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

GTACAAAATC GTTACAATAA AAC



## (2) INFORMATION FOR SEQ ID NO:6:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

CATCTTTAAT AATAGAAGAA TC

22

## Claims

1. An isolated polynucleotide comprising a polynucleotide sequence selected from the group consisting of:
  - (a) a polynucleotide having at least a 70% identity to a polynucleotide encoding a polypeptide comprising the amino acid sequence of SEQ ID NO:2;
  - (b) a polynucleotide having at least a 70% identity to a polynucleotide encoding the same mature polypeptide expressed by the *gidAl* gene contained in the *Chlamydia trachomatis* of the strain of the invention;
  - (c) a polynucleotide encoding a polypeptide comprising an amino acid sequence which is at least 70% identical to the amino acid sequence of SEQ ID NO:2;
  - (d) a polynucleotide which is complementary to the polynucleotide of (a), (b) or (c); and
  - (e) a polynucleotide comprising at least 15 sequential bases of the polynucleotide of (a), (b) or (c).
2. The polynucleotide of Claim 1 wherein the polynucleotide is DNA.
3. The polynucleotide of Claim 1 wherein the polynucleotide is RNA.
4. The polynucleotide of Claim 2 comprising the nucleic acid sequence set forth in SEQ ID NO:1.
5. The polynucleotide of Claim 2 comprising nucleotide 535 to 2364 set forth in SEQ ID NO:1.
6. The polynucleotide of Claim 2 which encodes a polypeptide comprising the amino acid sequence of SEQ ID NO:2.
7. A vector comprising the polynucleotide of Claim 1.
8. A host cell comprising the vector of Claim 7.
9. A process for producing a polypeptide comprising: expressing from the host cell of Claim 8 a polypeptide encoded by said DNA.
10. A process for producing a *gidAl* polypeptide or fragment comprising culturing a host of claim 8 under conditions

sufficient for the production of said polypeptide or fragment.

11. A polypeptide comprising an amino acid sequence which is at least 70% identical to the amino acid sequence of SEQ ID NO:2.

12. A polypeptide comprising an amino acid sequence as set forth in SEQ ID NO:2.

13. An antibody against the polypeptide of claim 11.

14. An antagonist which inhibits the activity or expression of the polypeptide of claim 11.

15. A method for the treatment of an individual in need of gidAl polypeptide comprising: administering to the individual a therapeutically effective amount of the polypeptide of claim 11.

16. A method for the treatment of an individual having need to inhibit gidAl polypeptide comprising: administering to the individual a therapeutically effective amount of the antagonist of Claim 14.

17. A process for diagnosing a disease related to expression or activity of the polypeptide of claim 11 in an individual comprising:

- (a) determining a nucleic acid sequence encoding said polypeptide, and/or
- (b) analyzing for the presence or amount of said polypeptide in a sample derived from the individual.

18. A method for identifying compounds which interact with and inhibit or activate an activity of the polypeptide of claim 11 comprising:

contacting a composition comprising the polypeptide with the compound to be screened under conditions to permit interaction between the compound and the polypeptide to assess the interaction of a compound, such interaction being associated with a second component capable of providing a detectable signal in response to the interaction of the polypeptide with the compound;  
and determining whether the compound interacts with and activates or inhibits an activity of the polypeptide by detecting the presence or absence of a signal generated from the interaction of the compound with the polypeptide.

19. A method for inducing an immunological response in a mammal which comprises inoculating the mammal with gidAl polypeptide of claim 11, or a fragment or variant thereof, adequate to produce antibody and/or T cell immune response to protect said animal from disease.

20. A method of inducing immunological response in a mammal which comprises delivering a nucleic acid vector to direct expression of gidAl polypeptide of claim 11, or fragment or a variant thereof, for expressing said gidAl polypeptide, or a fragment or a variant thereof *in vivo* in order to induce an immunological response to produce antibody and/ or T cell immune response to protect said animal from disease.